

ISYS1055/1057 Tutorial/Lab Sheet

Advanced Topics and Post-relational Databases

1. Suppose that we wish to build a genealogy database. We shall keep an entity set People – people are distinguished by social security No. (ssNo) and has a name. People are also involved in relationships Mother, Father and Children. Draw an ER diagram for this genealogy database, and map the ER diagram into a relational database schema.
2. Modify the genealogy database to include the special types of people --- Females, Males and People who are parents. Draw an ER diagram for the modified genealogy database and map the ER diagram into a relational database schema. Compare and discuss the resultant relations in relation to those of Question 1.
3. Files are stored as a sequence of fixed-size blocks on hard disks. A block (typically 512KB) is usually the unit for disk access (read and write). Disk access (read/write) is typically slow. Randomly fetching a block from disk takes about 10 milliseconds --- roughly the time required to execute 5, 000, 000 instructions in memory.
(Adapted from Example 6.44 of the main Text) Consider the relation:

StarsIn(movieTitle, movieYear, starName)

Assume that Data of StarsIn is kept in 100 disk blocks. Answer questions:

- 1) If there is not an index on the relation, how many block accesses are needed to find a tuple of the relation?
 - 2) Suppose an index is created on *movieTitle*, and the index uses a block. How many disk accesses are needed to find a tuple of the relation?
 - 3) Suppose that without using an index we can find a block on which an additional tuple fit, without scanning the entire relation. How many disk accesses are needed to insert a tuple?
4. Consider again the *StarsIn* relation. Assume that on average a star has appears in 2 movies and a movie has 4 stars. Suppose that there are 3 operations that we often perform on this relation, where S, T, and R are constants:
- Q1: SELECT movieTitle, movieYear
 FROM StarsIn
 WHERE starName=S;
- Q2: SELECT starName
 FROM StarsIn
 WHERE movieTitle=T and movieYear=R;
- I: INSERT INTO StarsIn VALUES (T, R, S);

Estimate the disk access cost for completing the above three queries assuming respectively that there is 1) not any index, 2) an index on *MovieTitle*, 3) an index on *StarName*, and 4) two indexes on *MovieTitle* and *StarName* respectively.